

# The *Amborella* pangenome suggests gene presence/absence variation is associated with environmental adaptation

Haifei Hu<sup>1</sup>, Armin Scheben<sup>1</sup>, Brent Verpaalen<sup>1</sup>, Soodeh Tirnaz<sup>1</sup>, Philipp E. Bayer<sup>1</sup>, Richard GJ Hodel<sup>2,3</sup>, Jacqueline Batley<sup>1</sup>, Douglas E Soltis<sup>2,3,4,5</sup>, Pamela S Soltis<sup>3,4,5</sup> and David Edwards<sup>1\*</sup>

<sup>1</sup>School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, WA, Australia

<sup>2</sup>Department of Biology, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup>Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

<sup>4</sup>The Genetics Institute, University of Florida, Gainesville, FL 32610, USA

<sup>5</sup>The Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA

Email: [\\*dave.edwards@uwa.edu.au](mailto:dave.edwards@uwa.edu.au)

## Abstract

*Amborella trichopoda*, an endemic flowering plant of New Caledonia, is the only living sister species of all other extant angiosperms. The pivotal phylogenetic position of *Amborella*<sup>1</sup>, make this a useful species for studying intraspecific genome variation in relation to local environmental adaptation. Here, we assembled the *Amborella* pangenome using an iterative mapping and assembly strategy based on ten individuals representing most of the known natural distribution. Dispensable genes are enriched in gene ontology terms associated with abiotic stress, suggesting that gene content may be driven by environmental adaption to specific regions on the island. Compared with crop pangenomes, the *Amborella* pangenome contains fewer disease resistance genes analogs (RGAs), and the RGAs it does contain are less likely to be dispensable. This pangenome for the sister species to all other extant angiosperms offers a chance to study pangenome structure and content in a non-crop plant species and provides a phylogenetic perspective for studying the evolution of PAVs.

## Results

An updated chromosome-level *Amborella* genome assembly was used as the starting reference. The assembled pangenome was 2.8% larger than the reference, and contained 30,189 genes, an increase of 9.2%. Modelling the total pangenome size suggested a closed pangenome with around 30,189 +/- 81 genes, of which, 27,440 +/- 566 were core genes found in all individuals (Figure 1). *Amborella* had relatively few dispensable genes (3,136, 10.4%). This contrasts with the gene PAVs for crop species such as *Brassica oleracea* (19%) and bread wheat (42%)<sup>2</sup>.

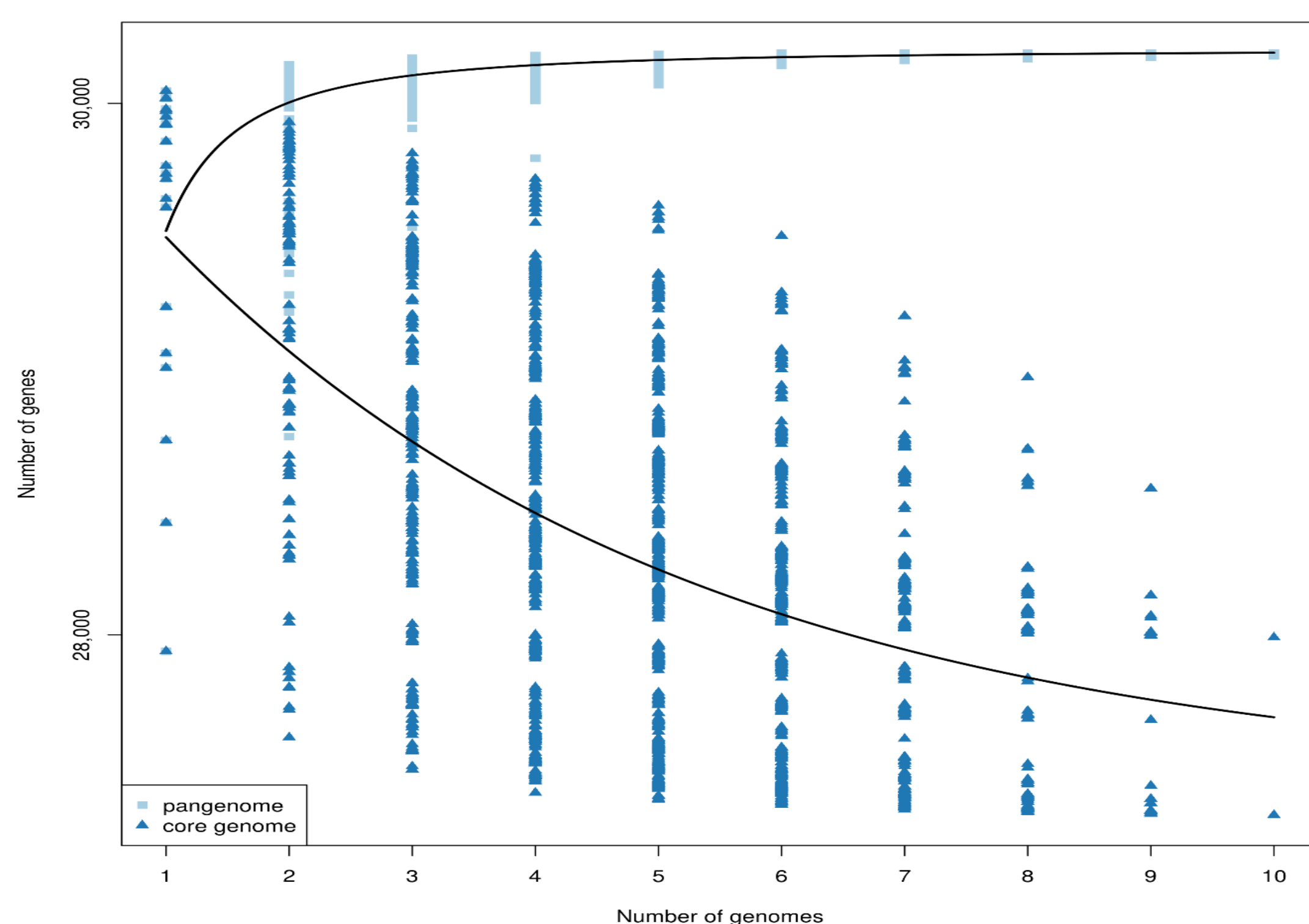


Figure 1: Simulations of the increase of the pan-genome size and the decrease of core-genome size.

GO enrichment analysis (Table 1) suggests that the dispensable genes were enriched for functions associated with environmental adaptation, predominantly responses to abiotic stress, including salt, cadmium and zinc, cold, water deprivation and heat. This contrasts with recent plant pangenome studies where the dispensable genes are predominantly enriched with biotic stress-related functions.

GO.ID	Term	Annotated pangenome genes	Annotated dispensable genes	p. value
GO:0046686	response to cadmium ion	320	143	< 1e-30
GO:0009651	response to salt stress	554	169	2.50E-24
GO:0009735	response to cytokinin	240	81	3.40E-21
GO:0009409	response to cold	320	112	2.70E-19
GO:0009414	response to water deprivation	346	99	2.50E-11
GO:0009408	response to heat	232	71	1.60E-10
GO:0009737	response to abscisic acid	555	121	3.50E-09
GO:0055114	oxidation-reduction process	2060	396	3.90E-09
GO:0009644	response to high light intensity	91	43	5.10E-09
GO:0042542	response to hydrogen peroxide	91	32	8.60E-09
GO:0009753	response to jasmonic acid	132	37	1.10E-07
GO:0010043	response to zinc ion	55	22	6.10E-07

Table 1: Significantly enriched gene ontology terms (with p value < 1E-07) among dispensable genes using all pangenome genes as a background.

One-third of New Caledonia is covered by mineral-rich soils. Variation in the amount and types of metal in the soil can give rise to diverse adaptation mechanisms. The geographic features and climate conditions vary across different locations, including a rain shadow, with 800 mm<sup>yr</sup><sup>-1</sup> in the western coastal region compared with 4,500 mm<sup>yr</sup><sup>-1</sup> on the eastern slopes of mountain areas<sup>3</sup>. The diverse environmental conditions present within a relatively small and isolated island may be a driver for the retention or loss of genes (Figure 2) for environmental adaptation across the island.

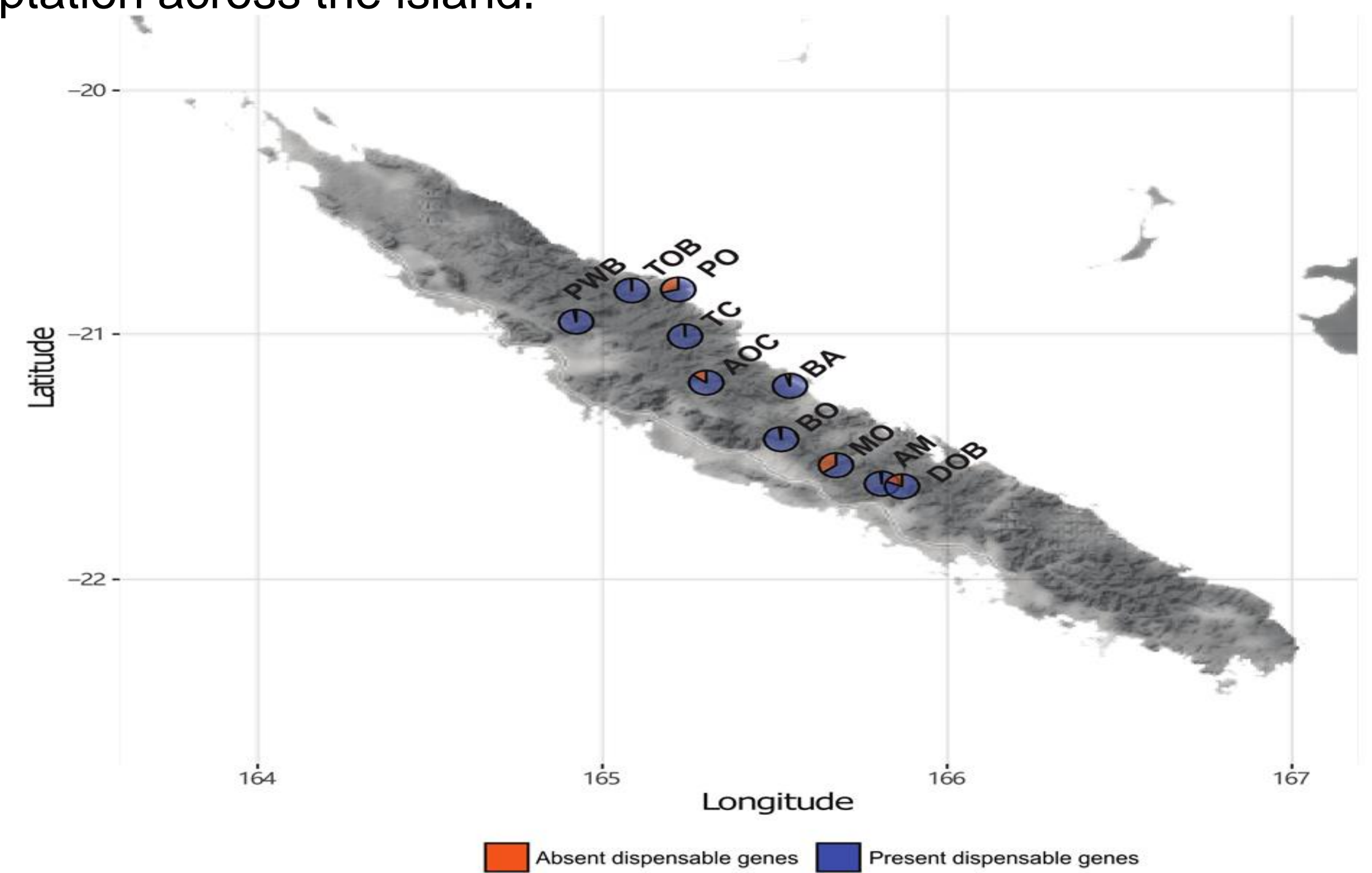


Figure 2: Percentage of present/absence dispensable genes in 10 different *Amborella* individuals growing in different geographical location in New Caledonia.

Further, we found that the *Amborella* pangenome contained relatively few (514) RGAs, compared to other crops. Only 4.7% RGAs were dispensable (Table 2). This contrasts with findings in other pangenome studies where at least 10 % of the RGAs were identified as dispensable genes<sup>4-6</sup>.

Speices	The total number of RGAs	The number of dispensable RGAs	Ref
<i>Amborella</i>	514	24 (4.7%)	
<i>Brassica oleracea</i>	1231	167(12%)	4
<i>Brassica napus</i>	1749	753(43%)	5
Soybean	2047	298(14.6%)	6

Table 2: Summary of the total number of RGAs and dispensable RGAs in different plant pangenome.

## Discussion

In this study, we constructed a pangenome and characterised genetic variation in the non-model flowering plant, *Amborella*. Fewer genes, especially RGAs, were dispensable in the *Amborella* pangenome compared to observations in crop species. This may be because *Amborella* has not experienced recent WGD, in contrast to many crop species. Furthermore, the species has undergone a series of bottlenecks during the past million years and occurs in geographically isolated populations, leading to limited interpopulational gene flow, both of which may reduce gene PAVs. Functional annotation of dispensable genes suggests that many of these genes are associated with abiotic stress, indicating that gene PAVs may be driven by environmental adaption to specific regions of New Caledonia.

## References and acknowledgements

1. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1241089 (2013).
2. Golicz, A.A., et al. Towards plant pangenomics. *Plant Biotechnology Journal* 14, 1099-1105 (2016).
3. Pouteau et al. New Caledonia: a Pleistocene refugium for rainforest lineages of relict angiosperms. *Journal of Biogeography* 42:2062–2077 (2015).
4. Bayer, Philipp E., et al. Variation in abundance of predicted resistance genes in the *Brassica oleracea* pangenome. *Plant biotechnology journal* 17.4: 789-800 (2018).
5. Dolatabadian, Aria, et al. Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant biotechnology journal* (2019).
6. Haifei Hu, et a;. Pangenome-wide study of RGAs in soybean. Unpublished manuscript.

We thank the rest of the members of the *Amborella* Genome Project for allowing us to use the updated and unpublished chromosome-level assembly in this analysis

